

7-2009

Evaluating Gender Bias in Ratings of University Instructors' Teaching Effectiveness

Suzanne Young

University of Wyoming, syoung@uwyo.edu

Leslie Rush

University of Wyoming, lrush@uwyo.edu

Dale Shaw

University of Northern Colorado, dale.shaw@unco.edu

Recommended Citation

Young, Suzanne; Rush, Leslie; and Shaw, Dale (2009) "Evaluating Gender Bias in Ratings of University Instructors' Teaching Effectiveness," *International Journal for the Scholarship of Teaching and Learning*: Vol. 3: No. 2, Article 19.

Available at: <https://doi.org/10.20429/ijstl.2009.030219>

Evaluating Gender Bias in Ratings of University Instructors' Teaching Effectiveness

Abstract

The purpose of this study was to examine the gender bias in student ratings of effective teaching. Students in five colleges were invited to rate instructors on three factors: interpersonal characteristics, pedagogical characteristics, and course content characteristics. We analyzed group differences based on student gender, instructor gender, and student level. Ratings of pedagogical characteristics and course content characteristics yielded significant interactions between student gender and instructor gender, but no differences were found among groups on interpersonal characteristics. We concluded that gender bias plays a role in students' views of effective teaching in terms of how students evaluate pedagogical and content characteristics and that this bias generalizes across student levels.

Keywords

University teaching effectiveness, Gender bias, Student ratings

Evaluating Gender Bias in Ratings of University Instructors' Teaching Effectiveness

Suzanne Young

University of Wyoming
Laramie, Wyoming, USA
syoung@uwyo.edu

Leslie Rush University
of Wyoming Laramie,
Woming, USA
Lrush@uwyo.edu

Dale Shaw

University of Northern Colorado
Greeley, Colorado, USA
dale.shaw@unco.edu

Abstract

The purpose of this study was to examine the gender bias in student ratings of effective teaching. Students in five colleges were invited to rate instructors on three factors: interpersonal characteristics, pedagogical characteristics, and course content characteristics. We analyzed group differences based on student gender, instructor gender, and student level. Ratings of pedagogical characteristics and course content characteristics yielded significant interactions between student gender and instructor gender, but no differences were found among groups on interpersonal characteristics. We concluded that gender bias plays a role in students' views of effective teaching in terms of how students evaluate pedagogical and content characteristics and that this bias generalizes across student levels.

Keywords: University teaching effectiveness, Gender bias, Student ratings

Introduction

How do students view effective teaching in higher education? Literally thousands of studies have addressed this issue, yet the question persists. As they have attempted to define good teaching, researchers have looked for differences in student evaluations based on type of course, class size, student abilities, and grading practices (Abrami, d'Apollonia, & Cohen, 1990; Greenwald & Gillmore, 1997). Researchers have also examined students' evaluations of teaching in terms of instructor and student characteristics, with inconsistent results (Basow, 2000; Hancock, Shannon, & Trentham, 1992; Marsh, 1987). The inconclusive nature of studies examining student gender, instructor gender, and student level, along with the emphasis on examining these characteristics individually, led us to focus on possible relationships among them. Drawing from Cohen's (1981) and Feldman's (1989) findings that students' evaluations of teaching are reliable and valid measures of good teaching, this study examines interactions among instructor, course, and student characteristics, and particularly focuses on instructor and student gender as well as student level.

Review of the Literature

Research on effective teaching suggests that student ratings are considered to be valid and reliable and are commonly used in university settings. Also, student ratings differ according to instructor characteristics, student characteristics, and course characteristics. The following provides a brief summary of some of the important findings relevant to this study.

Student Ratings of Teaching

Student ratings of teaching effectiveness have been shown to be valid measures of effective teaching. They are not only widely used in university settings but are also thoroughly reviewed in the literature. Cohen (1981), in a meta-analysis that examined the relationship between student ratings and student achievement, concluded that students are well equipped to rate their teachers when the criterion is student learning. Marsh (1987) reviewed student evaluation literature and, advocating the multi-method multi-trait technique to establish validity, found strong evidence of construct validity for the use of his instrument, Student Evaluation of Educational Quality (SEEQ). According to Greenwald and Gillmore (1997), validity of student ratings has been supported by reviews of research conducted since about 1980. Others (Feldman, 1988; Hativa, 1996; Murray, Rushton, & Paunonen, 1990) reported that student ratings were stable over time and consistent with ratings of others (peers, self-evaluations). Braskamp and Ory (1994) offered the opinion that "most faculty view student ratings as one important indicator of teaching ability," (p. 101) and that student ratings of teaching are both a valuable and credible source of information. The following sections examine research on student ratings of teaching effectiveness according to instructor characteristics, student characteristics, and course characteristics.

Instructor Characteristics

In relation to student ratings, instructor gender and other characteristics such as age, experience, and academic rank have been investigated extensively, in the United States and Canada, with mixed results. In examining the influence of instructor gender on student evaluations, for example, some researchers have found that female instructors are rated lower than their male colleagues (Basow & Silberg, 1987; Sandler, 1991); other researchers (e.g., Basow & Distenfeld, 1985; Feldman, 1983, 1993; Goodwin & Stevens, 1993; Hancock, Shannon, & Trentham, 1992) were unable to find evidence of gender differences. Still others, such as Feldman (2007), Bachen, McLoughlin, and Garcia (1999) and Tatro (1995) found that college students rated female instructors higher than male instructors.

Thus, it is probable that gender is a factor in students' evaluations of teaching, but that the relationship is a complex one (Basow, 2000). Students may associate certain types of behavior, such as teacher expressiveness, with gender; students' confusion of teaching styles and gender may also impact their evaluations (Arbuckle & Williams, 2003; Centra & Gaubatz, 2000). The setting in which such evaluations take place may also be important. Feldman, for example, conducted two reviews of literature examining how students rated male and female instructors in different ways. He found that very little gender bias was evident in classrooms in which extraneous variables were tightly controlled (Feldman, 1992), whereas a slight bias in favor of same gender preference took place in studies carried out in classrooms without such controls (Feldman, 1993). Arreola (2000), in a summary of studies on gender bias, suggests that the apparent bias may be due to courses that instructors are assigned to teach rather than the instructor's gender. In addition to the inconclusiveness of these and other gender studies, varied results with regard to instructor's age, experience, or academic rank are evident in the literature (Dukes & Victoria, 1989;

Renaud & Murray, 1996). From these studies, we may assume that, although there is some evidence that gender plays a role, there is still much research to do to better understand the impact of instructor gender, as well as other characteristics, on students' evaluations of teaching.

Student Characteristics

Student differences with regard to gender may contribute a great deal to the importance that students place on certain aspects of effective teaching. McKeachie (1990), in a commentary on research in college teaching, suggested that effective teaching is dependent on the characteristics of the students themselves, as well as on the teacher's behavior. Hancock, Shannon, and Trentham (1992), in a large study using students from five different colleges within the same university, found evidence that female students rated their teachers higher than male students on most aspects of effectiveness, except in the college of education. Tatro (1995), when asking both undergraduate and graduate students to evaluate their teachers, also found that female students rated teachers higher than did male students. Basow and Silberg (1987) found an interaction between student and instructor gender (males rated female teachers lower than male teachers and females rated male and female teachers very similarly) on most aspects of teaching effectiveness. Their sample, however, was limited to undergraduate students ($n=1029$) who may have been extremely traditional in gender roles, and they cautioned others not to interpret their results as strong evidence for gender differences. Bachen, McLouglin, and Garcia (1999) also found an interaction between student gender and instructor gender. In their study of approximately 500 university students' ratings, they found that female students rated female instructors higher on all five of their teaching dimensions: caring-expressive teaching style, professional-challenging, interactive, evaluation or feedback, and easy-going. However, male students did not view their male and female faculty differently on those same five factors. Summers, Anderson, Hines, Gelder, and Dean (1996) studied undergraduate and graduate students' perceptions of course satisfaction in traditional courses. They found that male students rated female instructors lower than did female students. Dukes and Victoria (1989), in a study using undergraduate students, found no significant differences among male and female ratings of teachers. These researchers call for additional study to identify what male and female students value in effective teachers.

Researchers have also examined the relationship of student age and student level with evaluations of teaching effectiveness. Basow and Silberg (1987) reported that there was a positive correlation between student level and teacher ratings for undergraduate students participating in their study. They assessed five factors: scholarship, organization and clarity, interaction with the group, interaction with individual students, and enthusiasm. Donaldson, Flannery, and Ross-Gordon (1993) reported comparative findings from three studies of adult students, concluding that adult graduate students identified some traits of effective teachers that were not typically mentioned by adult undergraduate students, such as clear presentation of material and teacher warmth. They also found that graduate students were more likely than undergraduate students to mention instructor characteristics such as role modeling, adaptation to student needs, providing motivation, using a variety of teaching techniques, openmindedness, and warmth as characteristics of effective instruction. In addition, they found developmental differences in age group expectations: younger students were most interested in attributes that might enhance their own tasks (that is, being successful in school) while older students were more attentive to relationship issues such as teachers who are dedicated and who motivate students to do their best. Donaldson et al. (1993) compared their findings to Feldman's (1988) meta-analysis of undergraduate students' views of effective college teachers and found that adult students

mentioned some characteristics that were not identified by Feldman. Adults, especially graduate students, appear to value dedicated teachers who create a comfortable learning atmosphere that is amenable to adaptation while they use a variety of teaching techniques.

Course Characteristics

Classroom or course characteristics such as class size, course discipline, course level, or whether a course was required or an elective have been found to relate to students' evaluations of teachers. Students in large classes generally tend to rate teachers lower than students in small classes (Feldman, 1984). Marsh (1987) and Marsh and Bailey (1993) found that graduate level courses were rated higher by students than undergraduate level courses. According to Feldman (1984), teachers delivering upper level courses have been consistently rated higher than those teaching lower level courses; elective courses receive higher marks than required courses; and the *soft* disciplines (for example, humanities and education) have higher rated teachers than the *hard* disciplines (such as mathematics and engineering). These findings hold little in the way of surprises but might be attributed to differences among students rather than differences in the effectiveness of teachers. In addition, Theall and Feldman (2007) suggest that researchers should consider conditions beyond the classroom itself such as online or distance education, private or for-profit institutions, impact of students' work and family responsibilities.

The purpose of the present study was to examine students' evaluations of teaching based on certain student and instructor characteristics. We expected that male and female students might differ on their ratings of instructors, depending on instructor gender. Also, we expected that we might find differences between undergraduate and graduate students' ratings. Previous studies, as shown by our review of literature, examined these characteristics without looking at interactions. Since we would be able to study interactions of student gender, student level, and instructor gender, we hypothesized that these interactions would help further our understanding of any possible gender bias and would clarify findings from previous studies.

Methodology

Undergraduate and graduate students ($n=765$) who were enrolled in a medium sized university in the western United States, were asked to participate in the study. The researchers found this sample by randomly selecting classes from the class schedule and stratifying by the five university's colleges; because of this approach, characteristics such as course type or student interest area would influence the findings only due to sampling error. Of the 40 instructors contacted, 34 agreed to give permission for the researchers to go into their classes and take no more than 15 minutes of class time to collect data. The six instructors who did not give permission declined for reasons such as a) they had activities planned that would require all of the class time, b) the class was not meeting during the time requested, or c) the class had been canceled. One of the six instructors did not respond to the request.

Using a twenty-five item instrument, students evaluated a memorable college or university teacher of their choice and not necessarily the instructor of the class the researchers were attending (see Appendix A). Students were asked to rate prior instructors in comparison to other university instructors they had encountered; rather than evaluating their courses (including course content and course delivery method), the students were asked to characterize the effectiveness of the teacher they chose to rate. This method focused

students' evaluation on teacher effectiveness, reducing the biases that may have existed due to their consideration of course grades. Prior to completing the evaluation instrument, the students participated in a brief discussion about rater errors, in an effort to raise their awareness level and decrease the effects of these errors. Specifically, the discussions addressed how biases can affect student ratings and students were asked to rate instructors carefully, honestly, and accurately. Students also discussed the scale, so that they would understand that they should compare the instructor to others they had known. There were two objectives for the discussion: first, that the errors associated with ratings might be reduced and, second, that the students would recognize that the purpose of the study was to understand effective teaching.

The twenty-five item instrument contained research-based items that had a demonstrated relationship with teacher effectiveness. Items included instructor subject matter knowledge, communication skills, concern for student learning, sense of humor, preparation for class, and others (Benz & Blatt, 1995; Feldman, 1988; Lowman, 1996; Marsh & Bailey, 1993). Because all of the items were literature based, content validity was strong. The items asked respondents to rate statements such as "The instructor was genuinely respectful of students" and "The instructor was knowledgeable about subject matter." All items were rated on a scale from one to nine, where one was *not at all descriptive* and nine was *very descriptive*.

We used factor analysis to reduce the twenty-five items for the sake of simplifying the interpretation. Using the maximum likelihood extraction method with a Promax rotation, three common factors were identified that accounted for 63.8% of the variance (see Table 1). A Promax rotation was used because the factors were assumed to be correlated and we were interested in interpreting the factors. The first factor, accounting for 55.4% of the variance and including 11 of the 25 items, primarily consisted of items that reflected how the instructors developed interpersonal relationships with students (interpersonal characteristics). The second factor was made up of eight items that were related to the instructors' teaching approaches (pedagogical characteristics) and accounted for 5.6% of the variance. The third factor, course content characteristics, was made up of four items that explained 2.8% of the variance. Only two items (appropriate assignments and appropriate evaluation methods) did not load on any of the three factors (loadings were less than .30). Since they were unique relative to the three common factors, they were not considered in further analyses.

Table 1. Factor loadings for 25 items on 3 factors

Factor number	Item Description	Factor 1	Factor 2	Factor 3
1	Warm and friendly	0.98	0.25	-0.01
1	Respect	0.91	-0.09	-0.02
1	Humor	0.79	0.01	0.02
1	Tolerance	0.79	-0.03	-0.04
1	Comfortable atmosphere	0.78	0.04	0.08
1	Adapt to student needs	0.74	0.10	0.01
1	Concern for student learning	0.61	0.23	0.06
1	Enjoyment	0.60	0.29	0.05

1	Enthusiasm	0.57	0.28	-0.30
1	Motivation	0.48	0.25	0.20
1	Accessible	0.36	0.14	0.04
2	Well prepared	-0.04	0.92	-0.16
2	Well organized	0.06	0.79	0.15
2	Clear explanations	0.09	0.76	0.06
2	Identify important ideas	-0.02	0.76	0.16
2	Subject matter knowledge	0.09	0.66	0.05
2	Use of good examples	0.08	0.65	0.17
2	Communication	0.28	0.55	0.06
2	Self-confident	0.05	0.55	0.01
3	Valuable course	0.06	0.04	0.94
3	Improved understanding	0.05	0.08	0.88
3	Increased interest	0.01	0.04	0.87
3	Worthwhile materials	0.01	0.19	0.49
(none)	Appropriate evaluation	0.24	0.27	0.26
(none)	Appropriate assignments	0.26	0.22	0.18

Note: Factor 1 is instructor characteristics, Factor 2 is pedagogical characteristics, and Factor 3 is course content characteristics.

Thus we grouped the twenty-five items from our instrument into three factors: interpersonal characteristics, pedagogical characteristics, and course content characteristics. Correlations among the three factors showed evidence of construct validity: $r = .56$ for instructor characteristics and pedagogical characteristics; $r = .24$ for instructor characteristics and course content characteristics; and $r = .37$ for pedagogical characteristics and course content characteristics. Since instructor and pedagogical characteristics tap into similar instructor traits, we expected the correlation between these two to be higher than correlations with course content characteristics. We then found the reliability for each factor; interpersonal characteristics, pedagogical characteristics, and course content characteristics were found to be .94, .93, and .91, respectively. These high reliabilities indicated that respondents were consistent in how they evaluated their instructors on these factors.

Next, we examined the means for interpersonal characteristics, pedagogical characteristics, and course content characteristics (our three factors) from several different perspectives. We began by simply examining student evaluations of instructors in terms of gender (both student and instructor) and student level (graduate or undergraduate). We then looked at interactions among student gender, instructor gender, and student level. We used Analysis of Variance (ANOVA) to compare these means. These results, including a description of the participants, are reported in the next section.

Results

The sample included 765 students who were enrolled in a variety of classes across the university. Of the 765 students, 246 were male and 519 were female; gender proportions were similar within the undergraduate and graduate groups (about one-third male and two-thirds female). The average age for the entire sample was 29.04; undergraduates reported an average age of 21.64 and graduates reported an average age of 34.93. Fifty-five percent ($n=424$) of the sample was graduate students and forty-five percent ($n=341$) was undergraduates. Also, most students (76.5%) chose to evaluate a course that was required for them. Students reported class sizes ranging from less than 10 to as many as 98 students (see table 2).

Table 2. Description of the sample

	n	Percent
Student Gender		
Male	246	32.2
Female	519	67.8
Instructor Gender		
Male	444	58.0
Female	321	42.0
Student Level		
Undergraduate	341	44.6
Graduate	424	55.4
Required course		
Yes	585	76.5
No	180	23.5
Class size		
Less than 20	146	19.1
20 to 39	413	54
40 to 59	120	15.7
Greater than 59	86	11.2

Three 3-way ANOVAs were conducted using the three factors (interpersonal characteristics, pedagogical characteristics, and course content characteristics) as dependent variables and student gender, instructor gender, and student level as independent variables. See Table 3 for the means of each of the three factors by independent variable.

Table 3. Means for the three factors by student gender, instructor gender, and student level

	Interpersonal characteristics	Pedagogical characteristics	Course content characteristics
Male instructors rated by male students	6.73	7.26	6.75
Female instructors rated by male students	6.62	6.69	6.23
Male instructors rated by female students	6.40	6.92	6.46
Female instructors rated by female students	6.77	7.28	6.78
Male instructors rated by all students	6.57	7.09	6.61
Female instructors rated by all students	6.69	6.99	6.51
Male students' ratings of all instructors	6.68	6.97	6.49
Female students' rating of all instructors	6.62	7.10	6.62
Undergraduate students' ratings of all instructors	6.62	7.03	6.41
Graduate students' ratings of all instructors	6.41	7.04	6.70
All students' ratings of all instructors	6.60	7.07	6.60

Note: Items were rated on a scale from one (not at all descriptive) to 9 (very descriptive). All items were written in a positive direction.

Tests of significance were conducted at the .05 level. Two of the three ANOVAs, analyzing differences in pedagogical characteristics and course content characteristics, yielded significant two-way interaction effects but no main effects (see Table 4) or three-way interactions. For interpersonal characteristics, groups did not differ significantly among any of the three independent variables or in their interactions.

Table 4. Analysis of variance for factors 1, 2, and 3

Source	F			p		
	F1	F2	F3	F1	F2	F3
Student Gender (A)	.37	.86	.53	.54	.35	.47
Student Level (B)	.02	.02	2.58	.88	.97	.11
Instructor Gender (C)	.66	.54	.31	.42	.46	.58
A x B	3.10	.00	.22	.08	.96	.64
A x C	2.52	11.88	5.37	.11	.01	.02
B x C	.04	.62	.02	.84	.43	.88
A x B x C	.19	.14	.04	.67	.71	.85

Note: F1=Factor 1 (interpersonal characteristics), F2=Factor 2 (pedagogical characteristics), F3=Factor 3 (course content characteristics). Significant effects are in bold type.

When we examined the differences among groups on pedagogical characteristics, we found a significant interaction between student gender and instructor gender ($p=.01$). Using LSD for follow-up comparisons, we found that male students rated male instructors ($M=7.26$) significantly higher than they rated their female instructors ($M=6.69$); female students rated female instructors ($M=7.28$) significantly higher than they rated their male instructors ($M=6.92$).

When we analyzed the third factor, course content characteristics, again we found a significant interaction between student gender and instructor gender ($p=.02$). This pattern was the same as that found for pedagogical characteristics, with male students rating male instructors ($M=6.75$) significantly higher than female instructors ($M=6.23$) and female students rating female instructors ($M=6.78$) significantly higher than male instructors ($M=6.46$).

In summary, male and female students (both undergraduate and graduate) rated their male and female instructors on three factors that related to effective teaching. The three factors were interpersonal characteristics, pedagogical characteristics, and course content characteristics. Female students rated their female instructors significantly higher on pedagogical characteristics and course content characteristics than they rated their male instructors. Also, male students rated male instructors significantly higher on the same two factors. Interpersonal characteristics of male and female instructors were not rated differently by the male and female students. Undergraduate and graduate students also did not rate their instructors differently and there was no interaction with the other two independent variables.

Discussion and Conclusions

The present study shows that student gender and instructor gender played an important role in how these students viewed good teaching. According to Centra and Gaubatz (2000),

a bias occurs when “a known characteristic of students systematically affects their ratings of teachers” (p. 17). In the present study, student gender interacted with instructor gender for two of the three factors of teaching effectiveness: pedagogical characteristics and course content characteristics. For both of these factors, students rated instructors of the same sex higher than instructors of the opposite sex. Thus, it is our perspective that gender bias played a role in student evaluations of instructors’ pedagogical characteristics and course content characteristics. That gender bias did not play a role in student evaluations of instructors’ interpersonal characteristics suggests that male and female students did not perceive a difference in their male and female instructors’ personality characteristics, such as warmth, friendliness, humor, and enthusiasm. This is in itself promising, as it shows that there is some potential for elimination of gender bias in students’ evaluations of their instructors. Additionally, students did not rate their instructors differently based on their level (undergraduate or graduate). Thus, we believe that the interaction of student gender and instructor gender, in other words gender bias, generalizes across student levels.

Our study sheds light on the mixed results of previous studies in regard to the impact of gender on differences in student ratings of their instructors. Because we used both instructor and student gender to examine groups of items on student ratings of their instructors, we developed a better understanding of exactly where gender bias plays a part in these ratings. Our findings show that in evaluation items related to pedagogical characteristics, such as organization, preparedness, and subject matter knowledge and to course content characteristics, such as perceived value of a course and student interest, gender bias is a potential complication of understanding and responding to student evaluations of instructors. Bachen, McLouglin, and Garcia (1999) also found an interaction between student and instructor gender; they found the same interaction pattern as we did in our study. It may be that male and female students may actually prefer different teaching styles and so they evaluate their male and female instructors differently.

Thus, the results of this study contribute to our understanding of the complexity of excellent teaching. The findings lend support to the views of Donaldson et al. (1993) and others (Feldman, 1988; Marsh, 1987; McKeachie, 1990; Young & Shaw, 1999) who purport that effective teaching is a complex construct imbedded within the context in which it takes place. Part of this context, of course, is the gender of both student and instructor. Although students in this study rated personal characteristics of their instructors in an unbiased manner, the ratings of items that were more closely aligned with content and with pedagogy showed gender bias. It is possible that expectations among students for how those pedagogical characteristics and perceptions of course content characteristics are experienced may differ depending on gender. For example, the markers for “organization” that may be important to a male student may differ from those that are important to a female student. Research in this area could help to ferret out some of these differences in expectations.

While the primary goal of this study is theoretical, there are obvious pragmatic applications for the findings. University supervisors and those responsible for the professional development of instructors can apply these findings several ways. First, it is important that awareness be raised of the potential for gender bias in ratings of pedagogical and course content characteristics for supervisors of instructors. Awareness of this tendency could result in a lowered dependence on the ratings in these areas, with a supplementation of other methods of evaluation for instructors. Secondly, instructor supervisors, along with instructors themselves, should work to make students aware of this potential for gender bias in evaluations and to help students become aware of their tendencies to allow this bias

to affect their ratings. Finally, instructors who have an awareness of this potential for bias in student evaluations could provide students with additional methods of giving feedback on the pedagogical techniques and course content of their courses. These additional methods of providing instructors with feedback could include discussions at the mid-term point or written reflections on evaluation criteria.

It is also important that future research in this area examine why students tend to show gender bias in their ratings. An exploratory study that asks students to reflect on and articulate the reasons for their rankings might be revelatory in understanding the depth of the gender bias, its source, and how it plays out in instructor rankings. This form of research might also be helpful in terms of developing and implementing evaluation tools, as well as articulating how best those tools might be utilized.

Previous research has led to an understanding that student ratings are a valid way to evaluate teaching, that students view the same teachers in different ways, that course and instructor characteristics are important, and that there are many ways in which teachers can be effective. The challenge for future research is to continue to study the complexities of effective teaching, including the effects of gender bias on students' evaluations of their instructors, so that evaluations can accurately reflect instructors' performance and so that instructors can use quality evaluations to improve their own teaching methods.

References

- Abrami, P. C., d'Apollonia, S. & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology*, 82(2), 219-231.
- Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles: A Journal of Research*, 49(9-10), 507-516.
- Arreola, R. A. (2000). *Developing a comprehensive faculty evaluation system* (2nd Ed.), Bolton, MA: Anker Publishing.
- Bachen, C. M., McLoughlin, M. M., & Garcia, S. (1999). Assessing the role of gender in college students' evaluation of faculty. *Communication Education*, 48, 193-210.
- Basow, S. A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles: A Journal of Research*, 34, 407-417.
- Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology*, 79(3), 308-314.
- Basow, S. A., & Distenfeld, M. S. (1985). Teacher expressiveness: More important for males than females? *Journal of Educational Psychology*, 77, 45-52.
- Benz, C., & Blatt, S. J. (1995). Factors underlying effective college teaching: What students tell us. *Mid-Westerner Educational Researcher*, 8(1), 27-31.

- Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work*. San Francisco: Jossey-Bass.
- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education*, 70(1), 17-30.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta analysis of multisection validity studies. *Review of Educational Research*, 51(3), 281-309.
- Donaldson, J. F., Flannery, D., & Ross-Gordon, J. (1993). A triangulated study comparing adult college students' perceptions of effective teaching with those of traditional students. *Continuing Higher Education Review*, 57(3), 147-163.
- Dukes, R. L., & Victoria, G. (1989). The effects of gender, status, and effective teaching on the evaluation of college instruction. *Teaching Sociology*, 17, 447-457.
- Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education*, 21, 45-116.
- Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities. *Research in Higher Education*, 28(4), 291-344.
- Feldman, K. A. (1989). The association between student ratings on specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 30, 137-194.
- Feldman, K. A. (1992). College students' views of male and female college teachers: Part I Evidence from the social laboratory and experiments. *Research in Higher Education*, 33(3), 317-351.
- Feldman, K. A. (1993). College students' views of male and female college teachers: Part II Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34(2), 151-211.
- Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. Perry & J. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93-129). Dordrecht, The Netherlands: Springer.
- Greenwald, A. G., & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89, 743-751.
- Goodwin, L. D., & Stevens, E. A. (1993). The influence of gender on university faculty members' perceptions of "good" teaching. *Journal of Higher Education*, 64(2), 166-185.
- Hancock, G. R., Shannon, D. M., & Trentham, L. I. (1992). Student and teacher gender in ratings of university faculty: Results from five colleges of study. *Journal of Personnel Evaluation in Education*, 6, 235-248.
- Hativa, N. (1996). University instructors' ratings profiles: Stability over time, and disciplinary differences. *Research in Higher Education*, 37(3), 341-365.

Lowman, J. (1996). Characteristics of exemplary teachers. *New Directions for Teaching and Learning*, 65, 33-40.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for further research. *International Journal of Educational Research*, 11, 253-388.

Marsh, H. W., & Bailey, M. (1993). Multidimensional students' evaluations of teaching effectiveness. *Journal of Higher Education*, 64(1), 1-18.

McKeachie, W. J. (1990). Research on college teaching: The historical background. *Journal of Educational Psychology*, 82(2), 189-200.

Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82(2), 250-261.

Renaud, R. D., & Murray, H. G. (1996). Aging, personality, and teaching effectiveness in academic psychologists. *Research in Higher Education*, 37(3), 323-340.

Sandler, B. R. (1991). Women faculty at work in the classroom, or, why it still hurts to be a woman in labor. *Communication Education*, 40, 6-15.

Summers, M., Anderson, J. L., Hines, A.R., Gelder, B. C., & Dean, R. S. (1996). The camera adds more than pounds: Gender differences in course satisfaction for campus and distance learning students. *Journal of Research and Development in Education*, 29(4), 212-218.

Tatro, C. N. (1995). Gender effects on student evaluations of faculty. *Journal of Research and Development in Education*, 28(3), 169-173.

Theall, M. & Feldman, K. A. (2007). Commentary and update on Feldman's (1997) "Identifying exemplary teachers and teaching: Evidence from student ratings." In R. Perry & J. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence based perspective* (pp. 130-143). Dordrecht, The Netherlands: Springer.

Young, S., & Shaw, D. G. (1999). Profiles of effective college and university teachers. *Journal of Higher Education*, 70, 670-686.

Appendix A

The teacher evaluation instrument

Teacher Evaluation Scale

Thank you for taking the time to answer the following questions about an instructor you have had in the recent past. As you rate your instructor within the context of a particular course, consider him/her relative to other university instructors you have had. Please rate each item indicating the degree to which you feel the item is descriptive of the instructor or course; where 1=not at all descriptive and 9=very descriptive. If you have no information or you feel the item does not apply, circle NA (Not applicable).

1. The instructor was knowledgeable about subject matter.
2. The instructor communicated effectively.
3. The instructor was enthusiastic about online teaching.
4. The instructor was well prepared for each class.
5. The instructor created a comfortable learning atmosphere.
6. The instructor adapted to student needs.
7. The instructor was tolerant of others' ideas and views.
8. The instructor was genuinely respectful of students.
9. The instructor was warm and friendly.
10. The instructor had a good sense of humor.
11. The instructor motivated students to do their best.
12. The instructor was self-confident.
13. The instructor genuinely enjoyed teaching.
14. The instructor was concerned about student learning.
15. The instructor was able to explain material clearly.
16. The instructor identified important ideas.
17. The instructor used good examples to explain concepts.
18. The instructor was accessible outside of class.
19. The assignments were appropriate in amount and level.
20. The evaluation methods were appropriate.
21. The course increased my interest in the subject matter.
22. The course was well organized.
23. The course materials (text, readings, etc.) were worthwhile.
24. The course improved my understanding of concepts in the field.
25. The course was valuable to me.

Please tell us a little about yourself and about the course.

You are: ☐ Male ☐ Female

Your age:

Your student level: ☐ Undergraduate ☐ Graduate

Approximate class size?

Your instructor was: ☐ Male ☐ Female

Was the course required? ☐ Yes ☐ No